New Links Between Differential and Linear Cryptanalysis

Céline Blondeau and Kaisa Nyberg

Aalto University, School of Science,
Department of Information and Computer Science
{celine.blondeau, kaisa.nyberg}@aalto.fi

Abstract. Recently, a number of relations have been established among previously known statistical attacks on block ciphers. Leander showed in 2011 that statistical saturation distinguishers are on average equivalent to multidimensional linear distinguishers. Further relations between these two types of distinguishers and the integral and zero-correlation distinguishers were established by Bogdanov et al. [6]. Knowledge about such relations is useful for classification of statistical attacks in order to determine those that give essentially complementary information about the security of block ciphers. The purpose of the work presented in this paper is to explore relations between differential and linear attacks. The mathematical link between linear and differential attacks was discovered by Chabaud and Vaudenay already in 1994, but it has never been used in practice. We will show how to use it for computing accurate estimates of truncated differential probabilities from accurate estimates of correlations of linear approximations. We demonstrate this method in practice and give the first instantiation of multiple differential cryptanalysis using the LLR statistical test on PRESENT. On a more theoretical side, we establish equivalence between a multidimensional linear distinguisher and a truncated differential distinguisher, and show that certain zero-correlation linear distinguishers exist if and only if certain impossible differentials exist.

Keywords: statistical cryptanalysis, block cipher, key-alternating block cipher, multiple differential attack, truncated differential, multidimensional linear attack, zero-correlation, impossible differential

1 Introduction

Block ciphers are used as building blocks for many symmetric cryptographic primitives for encryption, authentication, pseudo-random number generation, and hash functions. Security of these primitives is evaluated in regard to known attacks against block ciphers. Among the different types of attacks, the statistical ones exploit non-uniform behaviour of the data extracted from the cipher to find information about the secret key. Linear cryptanalysis [25] and differential cryptanalysis [4] are the most prominent statistical attacks against block ciphers.

Recently, a number of relations have been established among some previously known statistical attacks on block ciphers. Leander [23] observed that the

statistical saturation distinguishers [15] are on average equivalent to multidimensional linear distinguishers [20]. Further relations between these two types of distinguishers and the integral and zero-correlation distinguishers were established by Bogdanov et al. [6]. The goal of the work presented in this paper is to explore relations between linear and differential attacks. The strength of linear distinguishers relies on exceptionally high correlation, or a complete lack of it, while differential distinguishers are measured based on their probabilities. In the latter case also impossible differentials can be meaningful. The mathematical link between differential probability and linear correlation was presented by Chabaud and Vaudenay already in 1994 [12], but has never been used in practice due to its large computational complexity. In spite of this link, it is well known that resistance against differential cryptanalysis does not imply resistance against linear cryptanalysis. Also examples of the converse situation are known in the classical setting of distinguishers based on single differentials and single linear approximations [28]. In this paper, we will see that the situation changes when the distinguishers involve multiple differentials and linear approximations. Indeed, we will establish relations between multidimensional linear distinguishers and truncated differential distinguishers, and show, in particular, that existence of a zero-correlation relation is equivalent to existence of an impossible differential property.

The second major goal of the current paper is to apply the Chabaud-Vaudenay link in practice. The main motivation is due to the fact that, for some ciphers, it may be easier to evaluate probabilities of differentials than correlations of linear approximations, and for some other ciphers, the other way round. The block cipher PRESENT [5] is known to have a clear structure of linear approximations and their correlations have previously been evaluated accurately in [14] and [23]. On the other hand, the differentials over PRESENT split to numerous differential trails and their probabilities are hard to evaluate directly using traditional methods such as branch-and-bound algorithms [26, 7] or transition matrices [16].

As computation of the exact formula of the Chabaud-Vaudenay link between differential probabilities and squared correlations is not feasible, we develop a method based on theoretical arguments and assumptions to reduce the time complexity of the computation. The validity of these assumptions is then tested on a reduced-round version of PRESENT and PUFFIN [13].

Recently, an attack called multiple differential cryptanalysis (MDC) was proposed as an "all-in-one" generalisation of differential cryptanalysis [2, 9]. In these papers, distributions of differences for small block ciphers were evaluated to provide attacks using `LLR` and χ^2 scoring functions. This model, which improves and generalises differential, truncated differential, and impossible differential cryptanalysis methods remained, however, to be completed. To apply the `LLR` statistical test to actual block ciphers, cryptanalysts must be able to provide an upstream evaluation of the differential probabilities [9]. Up to now, computation of differential probabilities has been challenging for many ciphers. Given the method described above, we compute accurate estimates of truncated differen-

tial probabilities and give the first practical instantiation of multiple differential cryptanalysis using the LLR statistical test on PRESENT.

The rest of the paper is organised as follows. In Section 2, we first recall the basic definitions, the link between differential probabilities and linear correlations, and present the theoretical foundations for reducing the time complexity of using the link in practice. We then establish two new links between linear and differential cryptanalysis. The first one expresses the capacity of a multidimensional linear approximation in terms of a truncated differential probability, and the second one shows a relation between zero-correlation approximations and impossible differentials. In Section 3 we present the method for computing squared correlations for key-alternating block ciphers. Section 4 is devoted to the MDC method, the related LLR test, and its data complexity. In Section 5, parameters, like time complexity of the computation and time complexity of the MDC are described. In Section 6 we present the results from practical experiments and conclude in Section 7.

2 Links Between Differential and Linear Cryptanalysis

2.1 Differential Probabilities and Correlations of Linear Approximations

In differential cryptanalysis [4], the attacker is interested in finding and exploiting non-uniformity in occurrences of plaintext and ciphertext differences. Given a vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$, a differential is a pair (δ, Δ) where $\delta \in \mathbb{F}_2^n$ and $\Delta \in \mathbb{F}_2^m$ and its probability is defined as

$$\mathbf{P}[\delta \xrightarrow{F} \Delta] = \mathbf{P}_{\mathbf{X}}\left[F(\mathbf{X}) \oplus F(\mathbf{X} \oplus \delta) = \Delta\right],$$

where the probability is taken over the distribution of $\mathbf{X}$. Throughout this paper, it will be assumed that $\mathbf{X}$ is uniformly distributed in $\mathbb{F}_2^n$ in which case

$$\mathbf{P}[\delta \xrightarrow{F} \Delta] = 2^{-n}\#\{x \in \mathbb{F}_2^n \mid F(x) \oplus F(x \oplus \delta) = \Delta\}.$$

Linear cryptanalysis [25] uses a linear relation between bits from plaintexts, corresponding ciphertext and encryption key. Linear relations are expressed as Boolean functions of the plaintext and the key. The strength of the linear relation is measured by its correlation.

Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function. Its correlation is defined as its correlation with the all-zero function as

$$\mathbf{cor}_x(f) = 2^{-n}\Big[\#\left\{x \in \mathbb{F}_2^n | f(x) = 0\right\} - \#\left\{x \in \mathbb{F}_2^n | f(x) \neq 0\right\}\Big],$$

where the quantity within brackets can be computed as the Walsh transform of f evaluated at zero, see e.g. [11].

Given a vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ we are interested in Boolean functions $f(x) = a \cdot x \oplus b \cdot F(x)$ defined by linear relations where $a \in \mathbb{F}_2^n$ and

$b \in \mathbb{F}_2^m$ are called linear input and output masks. Chabaud and Vaudenay showed that differential probabilities and squared linear correlations are linked to each other by the Walsh transform .

Theorem 1 ([12]). *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a vectorial Boolean function. The probability of the differential (δ, Δ) over F can be given as*

$$\mathbf{P}[\delta \xrightarrow{F} \Delta] = 2^{-m} \sum_{a \in \mathbb{F}_2^n} \sum_{b \in \mathbb{F}_2^m} (-1)^{a \cdot \delta \oplus b \cdot \Delta} \mathbf{cor}_x^2 \left(a \cdot x \oplus b \cdot F(x)\right). \tag{1}$$

This formula has not been used before to compute differential probabilities of block cipher in practice. Indeed, the direct application of it would require computation and summing up 2^{n+m} squared correlations where n is the length of the input and m is the length of the output in bits of the function F. Later we will see that restricting attention to truncated differentials of a block cipher would allow us to reduce the size of the output space. Still, the problem with the large input space remains. Next, let us recall an important result of correlations of restrictions of Boolean functions.

Theorem 2. *Let $F : \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^m$ be a vectorial Boolean function, and let $x_t \in \mathbb{F}_2^t$ be uniformly distributed. Then*

$$\sum_{x_t \in \mathbb{F}_2^t} \mathbf{cor}_{x_s}^2 \left(a \cdot x_s + b \cdot F(x_s, x_t)\right) = 2^t \sum_{c \in \mathbb{F}_2^t} \mathbf{cor}_{x_s, x_t}^2 \left(a \cdot x_s + c \cdot x_t + b \cdot F(x_s, x_t)\right),$$

for all $a \in \mathbb{F}_2^s$ and all $b \in \mathbb{F}_2^m$.

This fact appeared in the context of Boolean functions as Lemma 4 of [24], see also [11], and was named as Fundamental Theorem in [27]. It describes the underlying principle for computing the average squared linear correlation, see Theorem 4 below, as well as for demonstrating the existence of the link between statistical saturation attack and the multidimensional attack [23]. We will use it now to derive the first result for reducing computations of differential probabilities according to Formula (1).

In our experiments, we observed that for SPN type block ciphers the number of active Sboxes at the first round influences the probability of the differential. Large probabilities can be found only with small number of active S-boxes. Hence, from the cryptanalyst's point of view, it seems reasonable to select the input difference δ to have a small Hamming weight. In such a situation we can apply Theorem 2 and reduce the space over which the correlations are computed.

Lemma 1. *Let $\mathbb{F}_2^n = \mathbb{F}_2^s \times \mathbb{F}_2^t$ and $\delta \in \mathbb{F}_2^n$ be such that $\delta = (\delta_s, \delta_t)$ where $\delta_s \in \mathbb{F}_2^s$ and $\delta_t \in \mathbb{F}_2^t$. If $\delta_t = 0$, then we have, for any fixed $b \in \mathbb{F}_2^m$,*

$$\begin{aligned} &\sum_{a \in \mathbb{F}_2^n} (-1)^{a \cdot \delta} \mathbf{cor}_x^2 \left(a \cdot x \oplus b \cdot F(x)\right) \\ &= 2^{-t} \sum_{x_t \in \mathbb{F}_2^t} \sum_{a_s \in \mathbb{F}_2^s} (-1)^{a_s \cdot \delta_s} \mathbf{cor}_{x_s}^2 (a_s \cdot x_s \oplus b \cdot F(x_s, x_t)), \end{aligned}$$

where $x = (x_s, x_t) \in \mathbb{F}_2^s \times \mathbb{F}_2^t$ and $a_s \in \mathbb{F}_2^s$.

This formula involves restricting the input space artificially by fixing a part of the input to $x_t \in \mathbb{F}_2^t$, and then taking the average over these fixations. According to our experiments this average can be accurately estimated in practice by restricting x_t to a small subset T of $\mathbb{F}_2^t$. How to choose T depends on the specific structure of the cipher under consideration and will help to reduce the time computation from 2^n to 2^s.

2.2 Links Between Multidimensional Linear Approximations and Truncated Differentials

In this section, we present new links between multidimensional linear and truncated differential attacks. A multidimensional linear relation (approximation) of a vectorial Boolean function is a linear space formed by a number of linear relations. Without loss of generality, we can assume that the input space and output space is split into two subspaces so that $F : \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^q \times \mathbb{F}_2^r$. Let us consider linear approximations of the form

$$(a_s, 0) \cdot x \oplus (b_q, 0) \cdot F(x),\ a_s \in \mathbb{F}_2^s,\ b_q \in \mathbb{F}_2^q,$$

and truncated differentials of the form

$$(\delta_s, *) \xrightarrow{F} (\Delta_q, *),\ \delta_s \in \mathbb{F}_2^s,\ \Delta_q \in \mathbb{F}_2^q,$$

and define the probability of such a truncated differential as

$$\mathbf{P}((\delta_s, *) \xrightarrow{F} (\Delta_q, *)) = 2^{-t} \sum_{\delta_t \in \mathbb{F}_2^t} \sum_{\Delta_r \in \mathbb{F}_2^r} \mathbf{Pr}((\delta_s, \delta_t) \xrightarrow{F} (\Delta_q, \Delta_r)).$$

Then by summing up on both sides of Equation (1) over all $\delta_t \in \mathbb{F}_2^t$ and $\Delta_r \in \mathbb{F}_2^r$, we obtain the following link between truncated differentials and multidimensional linear approximations.

Theorem 3. *For all $\delta_s \in \mathbb{F}_2^s$ and $\Delta_q \in \mathbb{F}_2^q$ it holds that*

$$\mathbf{P}((\delta_s, *) \xrightarrow{F} (\Delta_q, *)) = 2^{-q} \sum_{a_s, b_q} (-1)^{a_s \cdot \delta_s \oplus b_q \cdot \Delta_q} \mathbf{cor}_x^2((a_s, 0) \cdot x \oplus (b_q, 0) \cdot F(x)).$$

As an application of this result, let us consider a function, which satisfies an integral [17], for which some part of the output is uniformly distributed if some part of the input is fixed to an arbitrary value. One example of such a function is a three-round Feistel network with a bijective round-function. Another example is a function formed by three rounds backward or four rounds forward of the AES encryption function [22, 18]. As corollary of Theorem 3 we obtain the equivalence between such an integral condition and a condition on certain truncated differentials.

Corollary 1. *Let $F : \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^q \times \mathbb{F}_2^r$. Then the following are equivalent:*

(i) $\mathbf{cor}_{x_t}((b_q, 0) \cdot F(x_s, x_t)) = 0$ *for all* $x_s \in \mathbb{F}_2^s$ *and* $b_q \in \mathbb{F}_2^q \setminus \{0\}$,
(ii) $\mathbf{cor}_x((a_s, 0) \cdot x \oplus (b_q, 0) \cdot F(x)) = 0$ *for all* $a_s \in \mathbb{F}_2^s$ *and* $b_q \in \mathbb{F}_2^q \setminus \{0\}$,
(iii) $\mathbf{P}((\delta_s, *) \xrightarrow{F} (\Delta_q, *)) = 2^{-q}$ *for all* $\delta_s \in \mathbb{F}_2^s$ *and* $\Delta_q \in \mathbb{F}_2^q$,
(iv) $\mathbf{P}((0, *) \xrightarrow{F} (0, *)) = 2^{-q}$.

Proof. The equivalence of (i) and (ii) follows from Theorem 2. By Theorem 3, (ii) implies (iii). The implication from (iii) to (iv) is trivial, and finally, (iv) implies (ii) by Theorem 3.

The first condition means that the distribution of the first q bits of the output is uniform when taken over a fixed component x_s and variable component x_t in the input. Obviously, the conditions of Corollary 1 can hold only if $t \geq q$. In case $t = q$, we have the following equivalence between zero-correlation linear approximations and impossible differentials.

Corollary 2. *Let* $F : \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^t \times \mathbb{F}_2^r$ *be a vectorial Boolean function. Then all non-trivial linear relations* $(a_s, 0) \cdot x \oplus (b_q, 0) \cdot F(x)$, $a_s \in \mathbb{F}_2^s$, $b_q \in \mathbb{F}_2^q \setminus \{0\}$, *have correlation zero if and only if all non-trivial differentials* $(0, \delta_q) \xrightarrow{F} (0, \Delta_r)$, $\delta_q \in \mathbb{F}_2^q \setminus \{0\}$, $\Delta_r \in \mathbb{F}_2^r$, *are impossible.*

3 Key-Alternating Block Cipher

3.1 Linear Correlations

Let $\mathcal{E}_K : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a key-alternating block cipher, parametrised by a master key K, and comprising r' applications of the round function R_k, parametrised by the round key k. Let $(k_0, k_1, k_2, \cdots, k_{r'})$ be the round keys derived from the master key K. Without loss of generality, we assume that the key addition is the last component of the round function, that is, $R_{k_i}(x) = R(x) \oplus k_i$, for all $i = 1, \ldots, r'$. Then the block cipher E_K is defined as follows

$$\mathcal{E}_K(x) = R_{k_{r'}} \circ \cdots R_{k_2} \circ R_{k_1}(x \oplus k_0).$$

In the context of last rounds attacks, let us denote by F_K the first r rounds of the cipher. Then

$$\mathcal{E}_K(x) = R_{k_{r'}} \circ \cdots \circ R_{k_{r+1}} \circ F_K(x \oplus k_0).$$

By guessing (parts of) the keys $k_{r+1}, \ldots, k_{r'}$ the ciphertext can be (partially) decrypted over these rounds to achieve (partial) information about output data of F_K. Success of the attacks depends of many criteria. In the context of statistical attack, an evaluation of a non-uniform behaviour of r rounds of the cipher, allow the attacker to first build a distinguisher that will be used after to mount the attack.

As shown by Daemen [16] the correlation of a linear approximation $(a \cdot x \oplus b \cdot F_K(x))$ can be computed as a sum of key-dependent signed products of correlations of linear approximations that are chained over consecutive rounds. A chain of masks $U = (u_0, u_1 \cdots, u_r) \in (\mathbb{F}_2^n)^{r+1}$, where u_{i-1} and u_i are the input

and output masks over R at round i, is called a linear trail. If $k_0, \cdots, k_r$ are the round keys derived from a fixed master key K, then

$$\begin{aligned}&\mathbf{cor}_x\left(a \cdot x \oplus b \cdot F_{\mathrm{K}}(x)\right)\\&= \sum_{U; u_0=a; u_r=b} (-1)^{u_0 \cdot k_0 \oplus \cdots \oplus u_r \cdot k_r} \prod_{i=0}^{r-1} \mathbf{cor}_x(u_i \cdot x \oplus u_{i+1} \cdot R(x)). \qquad (2)\end{aligned}$$

Success and data complexity estimates in differential cryptanalysis are based on the average differential probabilities taken over all possible keys. We obtain a formula for this quantity by application of (1) for $F = F_K$, and then taking the average of both sides over K. It remains to compute the averages of the squared correlations. Next we recall the frequently used estimate of average squared correlations. This general form is obtained directly from Formula (2) by squaring both sides and taking the average over the round keys, or alternatively, by application of Theorem 2 by setting $y = K$ and $F(x, K) = F_K(x)$. By $\mathbf{E}_x(F(x))$ we denote the average value of F taken over x.

Theorem 4. *Using the notation given in this section and assuming that the round keys $k_0, \ldots, k_r$ are independent and uniformly distributed, we have*

$$\mathbf{E}_{k_0,\ldots,k_r}\left[\mathbf{cor}_x^2(a \cdot x \oplus b \cdot F_K(x))\right] = \sum_{\substack{U; u_0=u; \\ u_r=w}} \prod_{i=0}^{r-1} \mathbf{cor}_x^2(u_i \cdot x \oplus u_{i+1} \cdot R(x)). \quad (3)$$

3.2 Algorithm for Computing Average Squared Correlations

Daemen's formula (2) describes a way how to compute correlations round by round using correlation matrices. Similarly, Formula (3) can be implemented as a product of transition matrices corresponding to squared correlations of linear approximations over one round of the cipher.

In practice, as all correlations of one round linear approximations cannot be stored, a selection of the most significant linear approximations must be done, and only the squared correlations of the selected trails should be stored in the transition matrix. For instance, in the case of PRESENT, the single-bit linear trails are dominant, and a sharp estimate of the expected squared correlations of the cipher can be computed based only on these trails [14, 23].

Let Ω be a $N \times N$ matrix consisting of the squared correlations of the dominant one round linear approximations. We denote

$$\Omega[i, j] = \mathbf{E}_k\left[\mathbf{cor}_x^2(u_i \cdot x \oplus u_j \cdot R_k(x))\right],$$

where w_i, $i = 1, \ldots, N$, are the selected masks and k is the round key. Then by (3), if z rounds of the cipher with master key K is denoted by $R_{k^z}^{(z)}$, we have

$$\mathbf{E}_{k^z}\left[\mathbf{cor}_x^2(w_i \cdot x \oplus w_j \cdot R_{k^z}^{(z)})\right] \approx \Omega^z[i, j],$$

where Ω^z is the z-th power of the matrix Ω.

As only the masks corresponding to the most dominant approximations can be reached using the transition matrix Ω, rounds at the beginning and at the end should be added to complete the computation of the expected squared correlation for other input and output masks.

4 Multiple Differential Cryptanalysis

In the context of linear cryptanalysis, generalisations using distribution vectors and `LLR` and χ^2 statistical tests were provided first by Baignères, Junod and Vaudenay [3], and more recently, with applications to practical ciphers, by Hermelin, Cho, Nyberg [19]. For differential cryptanalysis, such multidimensional extensions appeared not until 2012 [2, 9]. In [2], a framework for such attacks was presented and tested for small block cipher. Cryptanalysis using multiple differentials on real ciphers, however, requires selection of suitable subsets of output differences, or grouping them in an appropriate way. In an attack model called *"unbalanced partitioning"* [9], a subspace of output differences is taken into consideration. In this model, the probability distributions involved ordinary differential probabilities, while the *"balanced partitioning"* model involves probability distributions of truncated differentials. The latter approach allows considering information from the whole output space. Advantages and disadvantages of both partitioning functions are discussed in [9]. In this article, we focus on MDC using balanced partitioning and probability distributions of truncated differentials, for the simple reason that those can be efficiently computed using the method of squared correlations.

4.1 Truncated Differentials

Let $F_K : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be, as before, r rounds of the block cipher. We aim at computing the probability distribution of truncated differentials, where the input difference δ is fixed, and the output differences are truncated and vary over all possible values. More concretely, let Δ be an output difference in a vector space $V \subset \mathbb{F}_2^n$. Let $\bar{V}$ be a complementary subspace of V, that is $\bar{V} \oplus V = \mathbb{F}_2^n$. Then $S_\Delta = \Delta \oplus \bar{V}$ is a truncated output difference and $\cup_{\Delta \in V} S_\Delta = \mathbb{F}_2^n$.

The probability of the truncated differential (δ, S_Δ) is defined[1]

$$\mathbf{P}\left[\delta \xrightarrow{F} S_\Delta\right] = \sum_{\gamma \in S_\Delta} \mathbf{P}\left[\delta \xrightarrow{F} \gamma\right] = \mathbf{P}\left[\delta \xrightarrow{G} \Delta\right], \tag{4}$$

where $G_K = \pi \circ F_K$ and π is a projection from $\mathbb{F}_2^n$ to V .
In what follows in this paper, we assume that δ is fixed and Δ takes all possible values in V. We study the non-uniformity of the distribution vector $p = [\mathbf{P}(\delta \xrightarrow{G} v)]_{v \in V}$, and denote $p_v = \mathbf{P}(\delta \xrightarrow{G} v)$, for $v \in V$. Then

$$p_v = \frac{1}{|V|} \cdot \sum_{a \in \mathbb{F}_2^n} \sum_{b \in V} (-1)^{a \cdot \delta \oplus b \cdot v} \mathbf{E}_K \left(\mathbf{cor}_x^2 \left(a \cdot x \oplus b \cdot G_K(x) \right) \right). \tag{5}$$

[1] A more general definition is given in Section 2.2

Using the optimisations given in Lemma 1 and Section 3.2 we can efficiently compute estimates of the expected values of the squared correlations $\mathbf{E}_K(\mathbf{cor}_x^2(a \cdot x \oplus b \cdot G_K(x))$, for all $a \in \mathbb{F}_2^n$ and $b \in V$. In all our experiments we compute correlations over two rounds "by hand" without the transition matrix at the beginning and at the end, to obtain the following formula

$$\frac{1}{|T|} \sum_{x_t \in T} \sum_{i,j=1}^{N} \mathbf{cor}_{x_s}^2(a_s \cdot x_s \oplus w_i \cdot R^2(x_s, x_t)) \Omega^{r-4}[i,j] \mathbf{cor}_x^2(w_j \cdot x \oplus b \cdot \pi(R^2(x)), \quad (6)$$

where in the computation of the first correlation $x = (x_s, x_t) \in \mathbb{F}_2^s \times \mathbb{F}_2^t$ and $a_s \in \mathbb{F}_2^s$. Sometimes, depending of the cipher more that two rounds of correlations can be used before going to selected correlations represented by the matrix Ω.

4.2 LLR Statistical Test and Data Complexity

We adopt the classical model of statistical cryptanalysis and assume that the *Wrong-Key Randomisation Hypothesis* holds. It means that for a wrong key guess the corresponding distribution is assumed to be uniform. We will denote the uniform distribution vector by $\theta = [\theta_v]_{v \in V}$ where each $\theta_v = \frac{1}{|V|}$.

When evaluating the security of the cipher or the complexity of a statistical distinguisher, accurate estimates of the differential probabilities are important. In [9], the authors studied the complexities of MDC for the LLR and the χ^2 distinguishers. When a good estimate of the expected probabilities is available, then the LLR distinguisher provides better data and memory complexities than the one based using χ^2 statistics. Nevertheless, it is well known that a small deviation in the estimation of the expected probability distribution will not allow the construction of a distinguisher using the LLR test. We demonstrate the accuracy of the estimates computed using Equation (6) by performing simulated attacks using the LLR distinguisher and by comparing the theoretical and the observed data complexity. Next we recall results from [3, 9] concerning the complexity of an attack using the LLR statistical test.

Definition 1. *Let $p = [p_v]_{v \in V}$ be the expected probability distribution vector, θ the uniform one and q^k the observed one for a key candidate k. For a given number of sample N_S, the optimal statistical test consists in comparing the following statistic to a fixed threshold.*

$$\mathtt{LLR}(q^k, p, \theta) \stackrel{\text{def}}{=} N_S \sum_{v \in V} q_v \log\left(\frac{p_v}{\theta_v}\right).$$

Definition 2. *Let p and p' be two probability distribution vectors over V. The* relative entropy *(aka.* Kullback-Leibler divergence*) between p and p' is*

$$D\left(p||p'\right) \stackrel{\text{def}}{=} \sum_{v \in V} p_v \log\left(\frac{p_v}{p'_v}\right).$$

We also define the following metrics

$$D_2\left(p||p'\right) \stackrel{\text{def}}{=} \sum_{v \in V} p_v \log^2\left(\frac{p_v}{p'_v}\right), \quad \textit{and} \quad \Delta D\left(p||p'\right) \stackrel{\text{def}}{=} D_2\left(p||p'\right) - D\left(p||p'\right)^2.$$

Theorem 5. *Let a be the advantage (see [31]) of an attack then the data complexity required to reach success probability P_S is*

$$N = 2 \cdot \frac{\left[\sqrt{\Delta D\left(p||\theta\right)}\,\Phi_{0,1}^{-1}(P_S) + \sqrt{\Delta D\left(\theta||p\right)}\,\Phi_{0,1}^{-1}\left(1 - 2^{-a}\right)\right]^2}{\left[D\left(p||\theta\right) + D\left(\theta||p\right)\right]^2}, \tag{7}$$

where $\Phi_{0,1}$ is the cumulative function of the standard normal distribution.

5 Practical Applications

Computation of the truncated differential probabilities using (6) depends on the ciphers. To compose the transition matrix, cryptanalyst must identify the important linear trails of the cipher. We consider this problem for two SPN block ciphers PRESENT[5] and PUFFIN[13].

5.1 Description of the Ciphers

The block cipher PRESENT is designed as a lightweight primitive which operates on 64-bit blocks of data. Ciphertexts are obtained after 31 iterations of the round function. The 16 Sboxes of PRESENT are all identical and are defined as a 4-bit non-linear permutation. PRESENT is parametrised by a 80-bit or a 128-bit key. More details on the specification can be found in [5].

The lightweight block cipher PUFFIN was introduced in [13]. It is defined as a 64-bit SPN block cipher parametrised with a 128-bit key. The round function as described in [13] is applied 32 times[2]. The structure of this cipher is similar to the one of PRESENT. By choosing involution components, the designers aim at efficient implementation in hardware.

Even if PUFFIN might not be of general interest, we selected it as a reference cipher for our experiments on PRESENT. As the Sboxes and the linear diffusion of these ciphers are essentially different, the linear and differential attacks have different impact on these ciphers. For PRESENT, linear cryptanalysis is more powerful (26 rounds [14]) than differential cryptanalysis (18 rounds [8, 34]). For PUFFIN, the best linear and differential types of attacks are about equally strong [10, 23]. The observed differences are largely due to the fact that PRESENT has the particularity of having strong single-bit linear relations over the S-box. The Sbox of PUFFIN is built in such a way that differences of Hamming weight one have high probabilities but the single-bit linear relations are not among the strongest. In our experiments, the transition matrix for PRESENT is composed

[2] Later, the same authors propose a new version of this cipher called PUFFIN2 [32].

of correlations for single-bit masks, while for PUFFIN we use a matrix consisting of all single-bit and two-bit linear approximations of the Sbox. By this choice we also aim at showing that the estimation method (6) works also in case when the single-bit linear trails are not dominant.

5.2 Parameters in Practice

In the context of differential cryptanalysis the attacker builds a distinguisher over all but a small number of the last rounds of the cipher and wants to recover information on the subkeys used at these last rounds. As partial decryption (inversion of some Sboxes) over the last rounds is time consuming, the ratio between the number of guessed keys and the number of Sbox inversions is often maximised. With this aim in mind, it is reasonable to choose a projected output space V which corresponds to a group of active Sboxes in the following round. Since PRESENT and PUFFIN use 4-bit Sboxes, we conduct experiments with $|V| = 2^4, 2^8, 2^{12}, 2^{16}$.

The MDC attack described in Section 4 takes into consideration all ciphertexts in the computation of the observed probability distributions. Contrary to classical differential cryptanalysis, there is no sieving, which means that the time complexity of the attack is always larger than the data complexity [9]. For instance, for an SPN cipher, where only part of the last round key is guessed during the attack, the time complexity is of the order of $|V| \times N$, where $|V|$ is the size of the projected output difference space and N the data complexity as derived in Theorem 5. As stated in [9], the memory complexity of multiple differential attacks using the `LLR` statistical test is dominated by the storage of the expected distribution p and the storage of an array of counters for recording the observed frequencies. When only the last round subkey is guessed, the memory complexity is then the storage of $2 \times |V|$ counters.

5.3 Time Complexity of Computation of Differential Probability

In practice, difficulty of the computation of the truncated differential probabilities using square correlations depends of the structure of the cipher and of the number of square correlations to compute. Formula (6) shows that for a fixed truncated differential, this computation can be decomposed into three steps consisting of computing the correlations over the first rounds, the intermediate rounds and the last rounds. In the case of PRESENT and PUFFIN an efficient computation of r-round squared correlation can be done using transitional matrices on $r-4$ rounds and by adding two rounds at the beginning and the end. For other ciphers than PRESENT, larger transition matrices should be taken into consideration. In the case of PUFFIN, computation of the powers of this matrix remain easy and fast using the two-bit linear trails. Using Lemma 1, computation for the first two rounds is done by computing the squared correlation over $x_s \in \mathbb{F}_2^s$ for a certain small number of restrictions specified by a set T of randomly selected x_t. Experiments show that the distribution of output differences is less uniform if the fixed input difference δ is selected such that

only one Sbox is active. Hence we can choose $x_s \in \mathbb{F}_2^4$. Computation over $T = 2^8$ random x_t has been seen to be enough for the ciphers studied in this paper. For the last two rounds, an average over 2^{20} random x, gives also a good estimate of the squared correlations $\mathbf{cor}_x(w_j \cdot x \oplus b \cdot \pi(R^2(x))$.

If we store values used many times, the time complexity of the computation of a truncated differential is dominated by the computation of the squared correlations over the first and and last rounds. It corresponds to a small number of encryptions. Using squared correlations and the transition matrix, computation of the expected differential probabilities can then be considered as independent of the number of rounds. In comparison, the complexity of the branch-and-bound algorithm is exponential in the number of rounds[8]. Hence, the computation of the truncated probabilities depends only on the size of V. For $V = 2^4$ this computation takes less than one minute on a standard computer.

In Section 4, we motivated why to use truncated differential probabilities in MDC. Truncated differential probabilities should be computed for all $v \in V$. As Formula (5) can be decomposed as

$$p_v = \frac{1}{|V|} \sum_{b \in V} (-1)^{b \cdot v} \cdot \sum_{a \in \mathbb{F}_2^n} (-1)^{a \cdot \delta} \mathbf{E}_K \left(\mathbf{cor}_x^2 \left(a \cdot x \oplus b \cdot G_K(x) \right) \right),$$

computation can be done efficiently by storing first the estimates of the sum over the input mask a computed using Lemma 1, for all $b \in V$. Then all p_v, $v \in V$, can be computed simultaneously using Fast Fourier Transform with time complexity $|V| \log |V|$.

6 Experiments and Attacks

6.1 Experiments

In this section we describe the experiments done with PRESENT and PUFFIN. We build an LLR distinguisher using the computed estimates of theoretical probability distributions over r rounds to attack $r+1$ rounds of the cipher. These experiments have been conducted in the following order: computation of the square correlation using transitional matrices, simulation of 100 multiple differential attacks using the `LLR` statistical test, and comparison between experimental data complexity and the theoretical one given by Theorem 5.

We conducted experiments on the ciphers PRESENT and PUFFIN with different numbers of rounds, different input differences, and different projected output spaces V of different sizes. Figure 1 illustrates the accuracy of the theoretical estimates in the case of the PRESENT cipher, for different sizes of V and different numbers of rounds. In this figure, we compare the differences in data complexity between the theoretical formula of Theorem 5 and the data requirements obtained using a mean over 100 simulated attacks. For the experiments presented in this figure, we selected $\delta =$ `0xf00000`. The advantage a is equal to 4 for $|V| \geq 2^8$, and equal to 2 for $|V| = 2^4$. The numbering of Sboxes corresponds to the one given in the specification [5]. Results of experiments on PUFFIN are

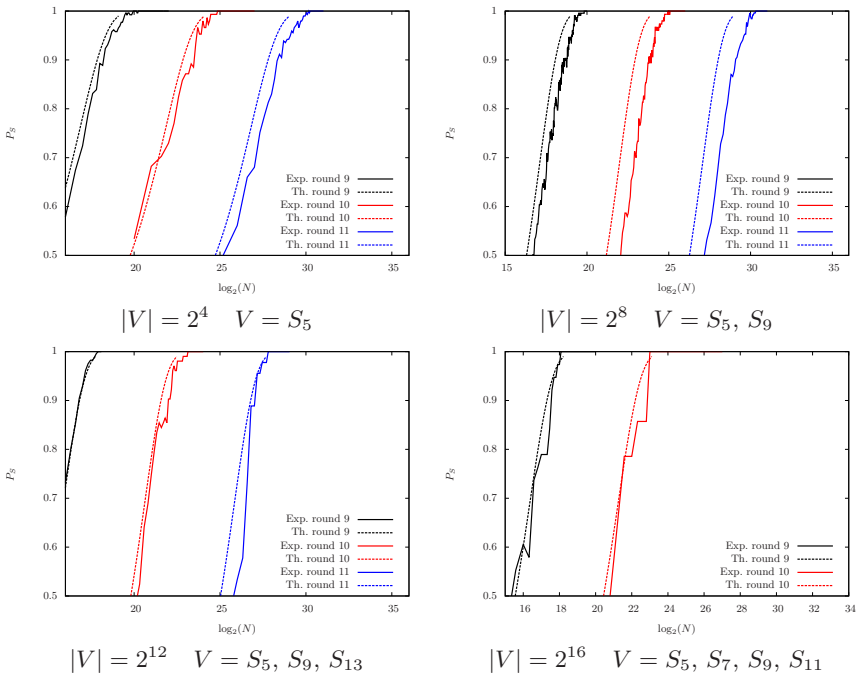


Fig. 1. Data complexity of attacks on 9, 10, 11 rounds of PRESENT

given in Appendix A. The obtained results from our simulations of MDC attacks are two-fold. First, it is well known that the `LLR` statistical test is efficient only if the analyst can provide a good estimate of the theoretical distribution. As the results of these experiments presented in Figure 1, 2 are tight, we can conclude that we were able to provide sufficiently accurate estimates of the differential probabilities using Formula (6). Secondly, we show for the first time an instantiation of an MDC attack on a full block size version of a state-of-the-art cipher.

6.2 Attacks on PRESENT

In the case of PRESENT with 80-bit key, the time complexity is bounded from above by 2^{80}. If the attack needs the full codebook then the size of the probability distribution must be less than 2^{16}. Different parameters are possible for the attack. As example, we propose an attack over 18 rounds using MDC distinguisher over 17 rounds. Parameters of this attack, with input difference $\delta =$ `0xf00000` and projected output difference space concentrated on Sboxes $S5$, $S9$ and $S13$, correspond to the ones used in attacks on a reduced-round version of the cipher (cf. Figure 1 with $|V| = 2^{12}$). Using the full codebook, this attack recovers 6-bits of the key with a success probability of 85% and has a time complexity of 2^{76} on which we add the exhaustive search of the 2^{74} remaining keys. Memory

complexity of this attack, is defined by the storage of the expected distribution vector and the storage of the counter array, and is equal to 2^{13} counters.
For the 128-bit key, partial inversion of the last two rounds is possible. Therefore using the same distinguisher over 17 rounds, we can propose an attack on 19 rounds. The choices of parameters for this attack are resumed in Table 1. We

Table 1. Parameters of attacks on PRESENT

#rounds	Key Length	Data Comp.	Adv.	Success Prob.	Time	Memory
18	80	2^{64}	6 bits	85%	$2^{76} + 2^{74}$	2^{13}
18	80	2^{62}	2 bits	85%	$2^{74} + 2^{78}$	2^{13}
19	128	2^{64}	6 bits	85%	$2^{124} + 2^{122}$	2^{60}
19	128	2^{62}	2 bits	85%	$2^{122} + 2^{126}$	2^{60}

conclude that the MDC distinguisher using truncated differentials described in this paper is the best distinguisher on PRESENT in the context of differential cryptanalysis. On the other hand, the key recovery attacks presented in this paper do not significantly improve over the previous differential attack on this cipher (18 rounds for both the 80-bit key and the 128-bit key). Best attacks on PRESENT are summarised in Table 2. We present a comparison between the attacks in this paper and the ones in [8, 34] which are based on simple differentials. Output differences of these simple differential focus on a small number of Sboxes. In this case, a sieving process can be applied, for both key schedules, and therefore one can invert two rounds of the cipher. Thus, using a 16-round distinguisher, 18 rounds can be attacked. In this paper, distribution of output differences over the the whole output space is taken into consideration. As no sieve is applied before guessing the key, the time complexity is larger and permits to invert only one round, for the 80-bit key. This explains why using the 17-round distinguisher, we are able to attack only 18 rounds of PRESENT-80 and 19 rounds for PRESENT-128.
Overall, the multiple differential attack presented in this paper corresponds quite well to the known differential properties of the PRESENT cipher. On the other hand, our simulations show that for PUFFIN truncated differentials do not provide better attacks than simple differential distribution.

7 Conclusion

Relations and dependencies between statistical attacks are of great importance when analysing the security of primitives based on block ciphers. In this paper, we extracted new relations between multiple differential and multidimensional linear distinguishers, and subsequently, between zero-correlation and impossible differential distinguishers. We used, for the first time, the relation between correlation of linear approximation and differential probability in practice to compute estimates of truncated differential probabilities of state-of-the-art ciphers from squared correlations of a selected set of linear approximations. We also derived a method to reduce the number of correlations needed to be computed, and in

this manner, succeeded to speed up the computation of these correlations to make the computation possible on a standard computer. Time complexity of this method is immune to the number of rounds, while for branch-and-bound algorithm it increases exponentially with the number of rounds.

The method developed in this paper was tested experimentally on the block ciphers PRESENT and PUFFIN and was further developed to a multiple differential attack on PRESENT which improves the best known attack in the differential context.

An interesting topic left for further research is to instantiate Theorem 3 on some ciphers and investigate it more closely. In this theorem, the truncated differentials and the multidimensional linear approximations occupy disjoint parts of the cipher, while in the method described in this paper the truncated differentials are located in the areas covered by known strong linear approximations. Therefore it may lead to essentially different results.

Table 2. Summary of the attacks on PRESENT.

#rounds	Version	Type of attack	Data	Time	Memory	Reference
16	80	Differential	$2^{64.0}$	$2^{64.0}$	$2^{32.0}$	[33]
18	80	Multiple Differential	$2^{64.0}$	$2^{64.0}$	$2^{32.0}$	[8]
18	80	Multiple Differential	$2^{64.0}$	$2^{64.0}$	$2^{32.0}$	[34]
18	**80**	**Multiple Differential (LLR)**	2^{62}	2^{78}	2^{13}	**This paper**
19	128	Algebraic Differential	$2^{62.0}$	$2^{113.0}$	n/r	[1]
19	**128**	**Multiple Differential (LLR)**	2^{62}	2^{126}	2^{60}	**This paper**
24	80	Linear	$2^{63.5}$	$2^{40.0}$	$2^{40.0}$	[30]
24	80	Statistical Saturation	$2^{57.0}$	$2^{57.0}$	$2^{32.0}$	[15]
25	128	Linear	$2^{64.0}$	$2^{96.7}$	$2^{40.0}$	[29]
26	80	Multiple Linear	$2^{64.0}$	$2^{72.0}$	$2^{32.0}$	[14]

Acknowledgments. We wish to thank Gregor Leander for useful discussions that led us to discovery of the results presented in Section 2.2. We also wish to thank the anonymous reviewers for helpful comments.

References

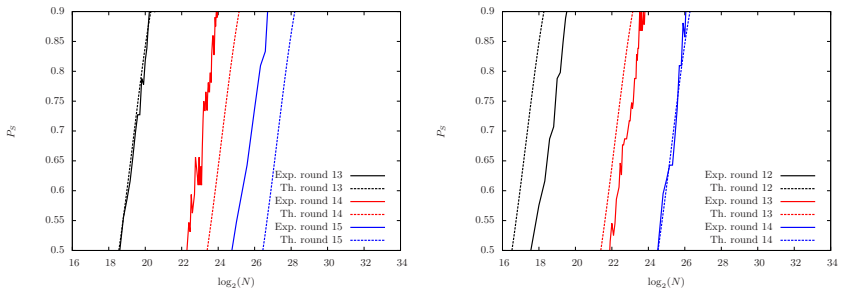
1. Albrecht, M., Cid, C.: Algebraic techniques in differential cryptanalysis. In Dunkelman, O. ed.: FSE 2009. Vol. 5665 of LNCS., Springer (2009) 193–208
2. Albrecht, M. R. and Leander, G.: An All-In-One Approach to Differential Cryptanalysis for Small Block Ciphers. In Knudsen, L. R. and Wu, K. eds.: SAC 2012, Vol. 7707 of LNCS, Springer (2012) 1–15.
3. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In Lee, P.J. ed.: ASIACRYPT 2004. Vol. 3329 of LNCS., Springer (2004) 432–450
4. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. In Menezes, A., Vanstone, S.A., eds.: CRYPTO 1990. Vol. 537 of LNCS., Springer (1991) 2–21

5. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: An ultra-lightweight block cipher. In Paillier, P., Verbauwhede, I., eds.: CHES 2007. Vol. 4727 of LNCS., Springer (2007) 450–466
6. Bogdanov, A., Leander, G., Nyberg, K., Wang.M.: Integral and Multidimensional Linear Distinguishers with Correlation Zero. In Wang, K.,Sako, K., eds : ASIACRYPT 2012 Vol. 7658 of LNCS., Springer (2012) 244-261
7. Blondeau, C., Gérard, B.: Links between theoretical and effective differential probabilities: Experiments on PRESENT. In: TOOLS'10. (2010)
8. Blondeau, C., Gérard, B.: Multiple Differential Cryptanalysis: Theory and Practice. In Joux, A., eds.: FSE 2011. Vol 6733 of LNCS, Springer (2011) 35 – 54
9. Blondeau, C., Gérard, B., Nyberg, K.: Multiple Differential Cryptanalysis Using `LLR` and χ^2 Statistics In Visconti, I., De Prisco, R. eds: SCN 2012, Vol. 7485 of LNCS, Springer (2012) 343–360
10. Blondeau, C., Gérard, B.: Differential Cryptanalysis of PUFFIN and PUFFIN2. In ECRYPT Workshop on Lightweight Cryptography - LC 2011. (2011).
11. Carlet, C.: Boolean Functions for Cryptography and Error Correcting Codes. In Cambridge University Press (To Appear)
12. Chabaud, F., Vaudenay, S.: Links Between Differential and Linear Cryptoanalysis. In De Santis, A. ed: EUROCRYPT-94, Vol. 950 of LNCS, Springer (1994), 356–365
13. Cheng, H., Heys, .M., Wang, C.: "PUFFIN: A Novel Compact Block Cipher Targeted to Embedded". In Fanucci, L., ed.: DSD 2008, IEEE (2008) 383–390
14. Cho, J.Y.: Linear cryptanalysis of reduced-round PRESENT. In Pieprzyk, J., ed.: CT-RSA 2010. Vol. 5985 of LNCS, Springer (2010) 302–317
15. Collard, B., Standaert, F.X.: A statistical saturation attack against the block cipher PRESENT. In Fischlin, M. ed.: CT-RSA 2009. Vol. 5473 of LNCS, Springer (2009) 195–210
16. Daemen, J., Govaerts, R., Vandewalle, J. Correlation matrices. In Preneel, B. ed.: FSE 1994. Vol. 1008 of LNCS, Springer (1995) 275–285
17. , Daemen J., Knudsen, L., Rijmen, V. The Block Cipher Square, In Biham, E. ed: FSE 1997. Vol. 1267 of LNCS, Springer (1997) 149–165
18. Gilbert, H.: An untwisted representation of AES, Early Symmetric Crypto seminar, Mondorf-les-Bains, Luxemburg, January 2013
19. Hermelin, M., Cho J. Y., Nyberg, K.: A new technique for multidimensional linear cryptanalysis with applications on reduced round Serpent. In Lee, P.J., and Cheon, J.H. eds: ICISC 2008, Vol. 5461 of LNCS, Springer (2009) 383–398
20. Hermelin, M., Cho J. Y., and Nyberg, K.: Multidimensional extension of Matsui's algorithm 2. In Dunkelman, O. ed.: FSE 2009, Vol. 5665 of LNCS, Springer (2009) 209–227
21. Knudsen, L.R.: Truncated and Higher Order Differentials. In Preneel, B. ed.: FSE 1994. Vol. 1008 of LNCS, Springer (1995) 196–211
22. Knudsen, L.R., Rijmen, V.: Known-key distinguishers for some block ciphers. In: Kurosawa, K. ed.: ASIACRYPT 2007, Vol. 4833 of LNCS, Springer (2007) 315–324
23. Leander, G.: On Linear Hulls, Statistical Saturation Attacks, PRESENT and a Cryptanalysis of PUFFIN. In Paterson, K.G. ed.: EUROCRYPT 2011. Vol. 6632 of LNCS, Springer (2011) 303–322
24. Linial, N., Mansour, Y., and Nisan, N.: Constant Depth Circuits, Fourier Transform, and Learnability. Journal of the Association for Computing Machinery, Volume 40, No 3, July 1993, 607–620
25. Matsui, M.: Linear cryptanalysis method for DES cipher. EUROCRYPT '93, Vol. 765 of LNCS, Springer (1993) 386–397

26. Matsui, M.: On Correlation Between the Order of S-boxes and the Strength of DES. EUROCRYPT '94, Vol. 950 of LNCS., Springer (1994) 366–375
27. Nyberg, K.: Linear approximation of block ciphers. In De Santis, A. ed: EUROCRYPT'94, Vol. 950 of LNCS, Springer (1995) 439–444.
28. Nyberg, K.: S-boxes and round functions with controllable linearity and differential uniformity. In Preneel, B. ed: FSE 1994 Vol. 1008 of LNCS, Springer(1995) 111–130
29. Nakahara, J., Sepehrdad, P., Zhang, B., Wang, M.: Linear (hull) and algebraic cryptanalysis of the block cipher PRESENT. In Garay, J. A., Miyaji, A., Otsuka, A. eds: CANS 2009. Vol. 5888 of LNCS, Springer (2009) 58–75
30. Ohkuma, K.: Weak keys of reduced-round PRESENT for linear cryptanalysis. In Jacobson, M. J., Rijmen, V., Safavi-Naini, R. eds.: SAC 2009. Vol. 5867 of LNCS, Springer (2009) 249–265
31. Selçuk, A. A.: On Probability of Success in Linear and Differential Cryptanalysis. Journal of Cryptology. Vol. 21 Number 1 (2008) 131–147
32. Wang, C., Heys, H. M.: An ultra compact block cipher for serialized architecture implementations. In Fanucci, L. ed.: CCECE 2009, IEEE (2009) 1085–1090
33. Wang, M.: Differential cryptanalysis of reduced-round PRESENT. In Vaudenay, S. ed.: AFRICACRYPT 2008. Vol. 5023 of LNCS, Springer (2008) 40–49
34. Wang, M., Sun, Y., Tischhauser,E., Preneel, B.: A Model for Structure Attacks, with Applications to PRESENT and Serpent. In Canteaut, A. ed.: FSE 2012. Vol. 7549 of LNCS, Springer (2012) 49–68

A Appendices

Figure 2 presents results of some experiments on a reduced-round version of PUFFIN. Distribution over r rounds was computed using Formula (6) to simulate multiple differential attack on $r+1$ rounds. Different experiments have been conducted. In this figure we illustrate the results for the input difference $\delta = \texttt{0x2}$ and output projected space V concentrated on $S3$ and $S5$ (for $|V| = 2^8$) and $S3$, $S5$, $S9$ (for $|V| = 2^{12}$).



$|V| = 2^8$ $|V| = 2^{12}$

Fig. 2. Data complexity of attacks on 12, 13, 14 rounds of PUFFIN